

» 巨量資料分析與應用 (2)

楊立偉教授

台灣科技大學資管系

2013 Fall

楊立偉教授

◆ 專長領域為資料庫及語意分析技術、知識管理、數位行銷

現任	台科大資管系兼任助理教授 2008~
	台大工管系暨商研所兼任助理教授 2006~
	資訊及通信國家標準技術委員
	意藍資訊 董事總經理（創辦人） 1999~ 國內規模最大的網路情報與社群口碑自動分析平台
	龍捲風科技 董事總經理 國內企業搜尋引擎市佔率最高；國際檢索競賽第一名
經歷	智威湯遜數位行銷首席顧問、尚藍互動行銷共同創辦人
	2009年獲選100 MVP最有價值經理人，擁有超過20項語意分析專利
	2012年榮獲國家雲端創新獎、數位時代「創業之星」首獎

課程大綱

◆ 第一部份

- 巨量資料導論
- 巨量資料分析與管理架構
- 巨量資料分析技術

◆ 第二部份

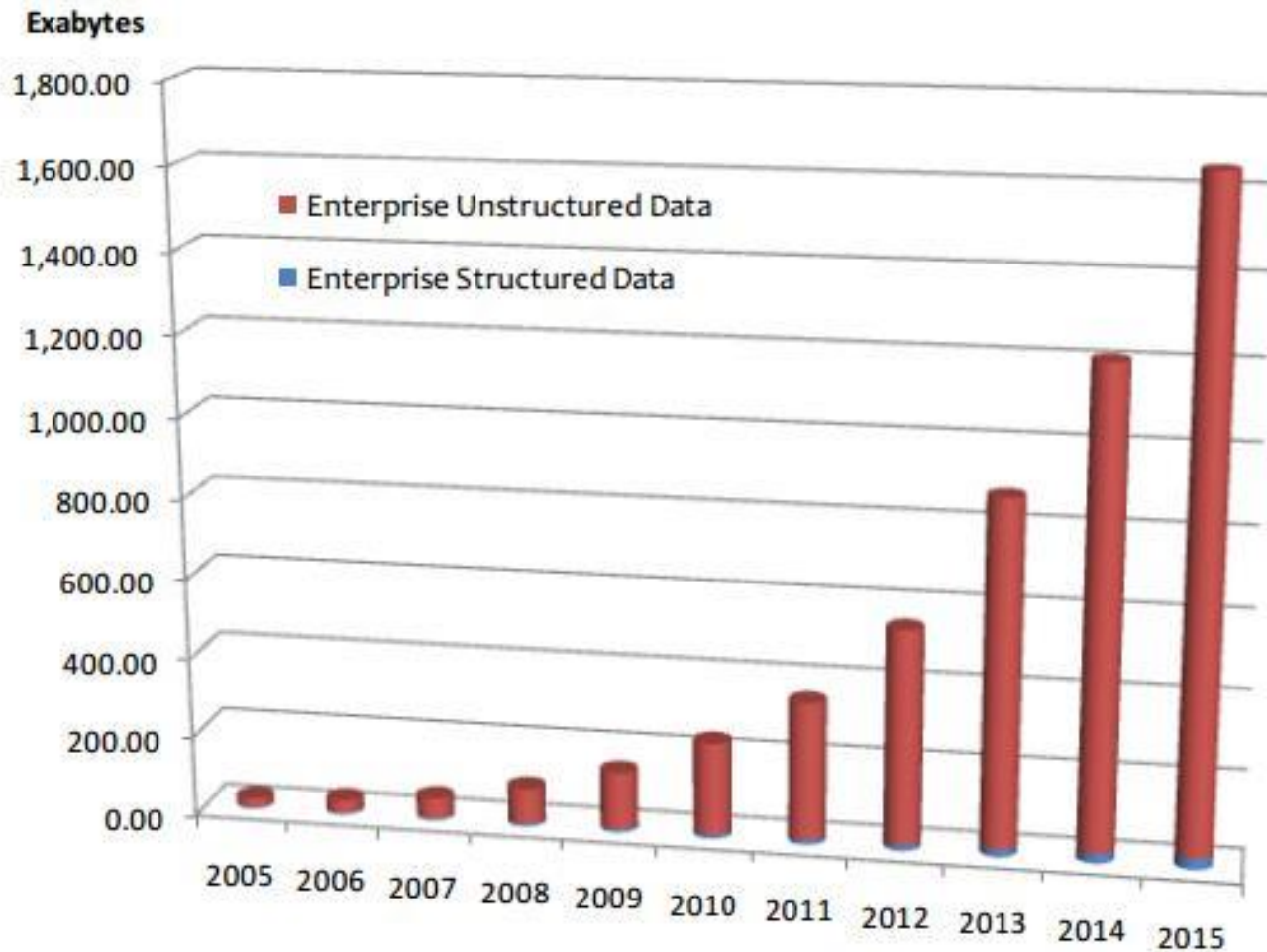
- 應用案例與研討 – 企業個案 (1)
- 應用案例與研討 – 企業個案 (2)
- 應用案例與研討 – Open Data



巨量資料分析技術 (2)

Big data 的資料種類

◆ 企業的結構性資料 與 非結構性資料



◆ Twitter

- 200 million tweets per day
- Peak 10,000 per second
- How to analyze the data ?



◆ Zynga

- "Analytics company, not a gaming company "
- 230 million players per month
- Harvest 15TB data per day
 - test new features
 - target advertising

4U box = 40 TB
1PB = 25 boxes



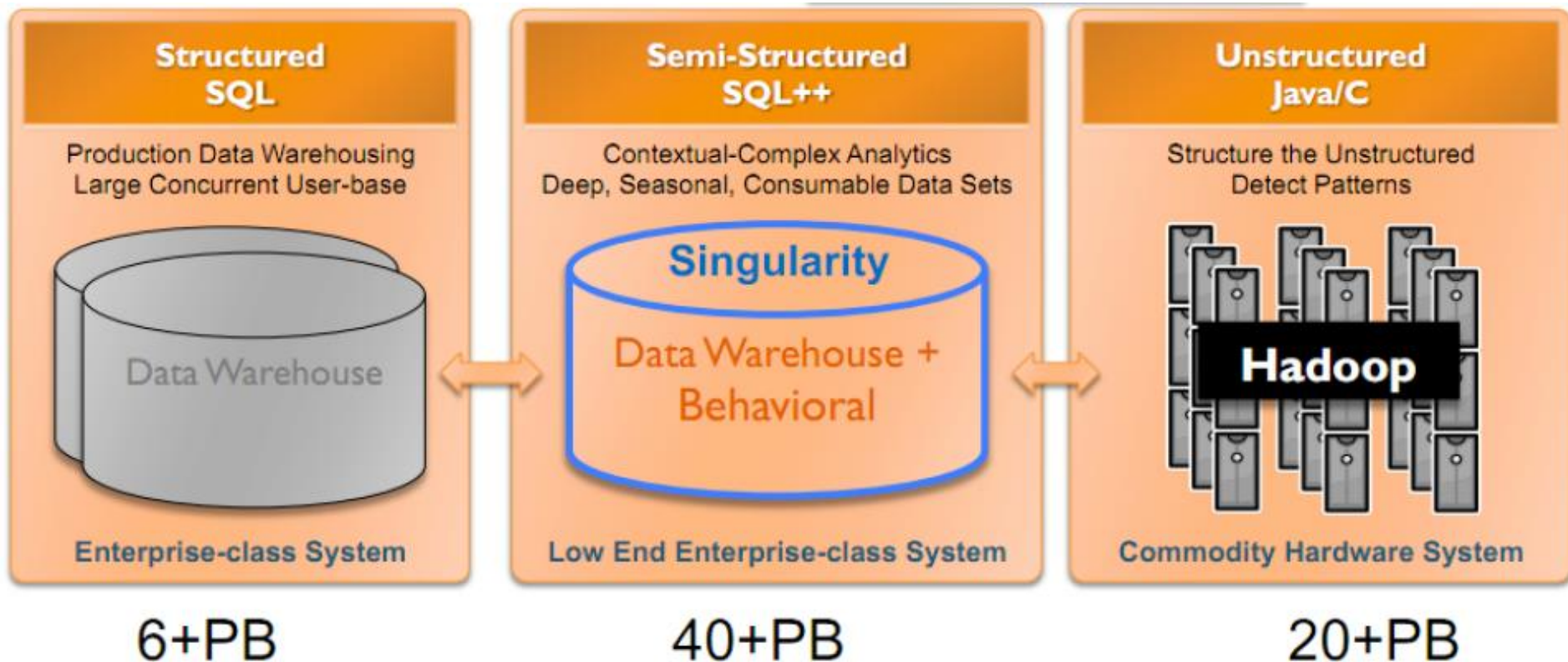
◆ Facebook

- 6 billion messages per day
- 2 PB (compressed) online
- 6 PB replication
- 250 TB growth per month
- Cassandra / HBase architecture

The Facebook logo, consisting of the word "facebook" in white lowercase letters on a dark blue rectangular background.

Analyze & Report

Discover & Explore



Big data 的分析方式

- ◆ 結構性資料分析
 - Data Mining 資料探覬
- ◆ 非結構性資料分析
 - Text Mining 文字探覬
 - 轉結構性資料

(1) 結構性資料分析

◆ 常見的 Data Mining 模組

- 群集分析 Clustering
- 分類預測 Classification
- 關聯規則 Association rules
- 連續行為 Sequential pattern analysis

◆ 資料探戡 Data Mining

- 從大量資料或歷史資料中，透過人工智慧、類神經網路等技術，找出人類難以觀察出、或傳統之統計所無法發現之隱性知識。
- 這些隱性知識包括分布趨勢、決策樹、關聯性、連續性等，對行銷與客戶關係經營有重大價值。

基本原理：以相關性分析為例

Database 結帳紀錄

TID	Itemset
100	1 3 4
200	2 3 5
300	1 2 3 5
400	2 5



L₁

Itemset	Support
{1}	2
{2}	3
{3}	3
{4}	1
{5}	3



L₂

Itemset	Support
{1 3}	2
{1,4}	1
{2 3}	2
{2 5}	3
{3,4}	1
{3 5}	2



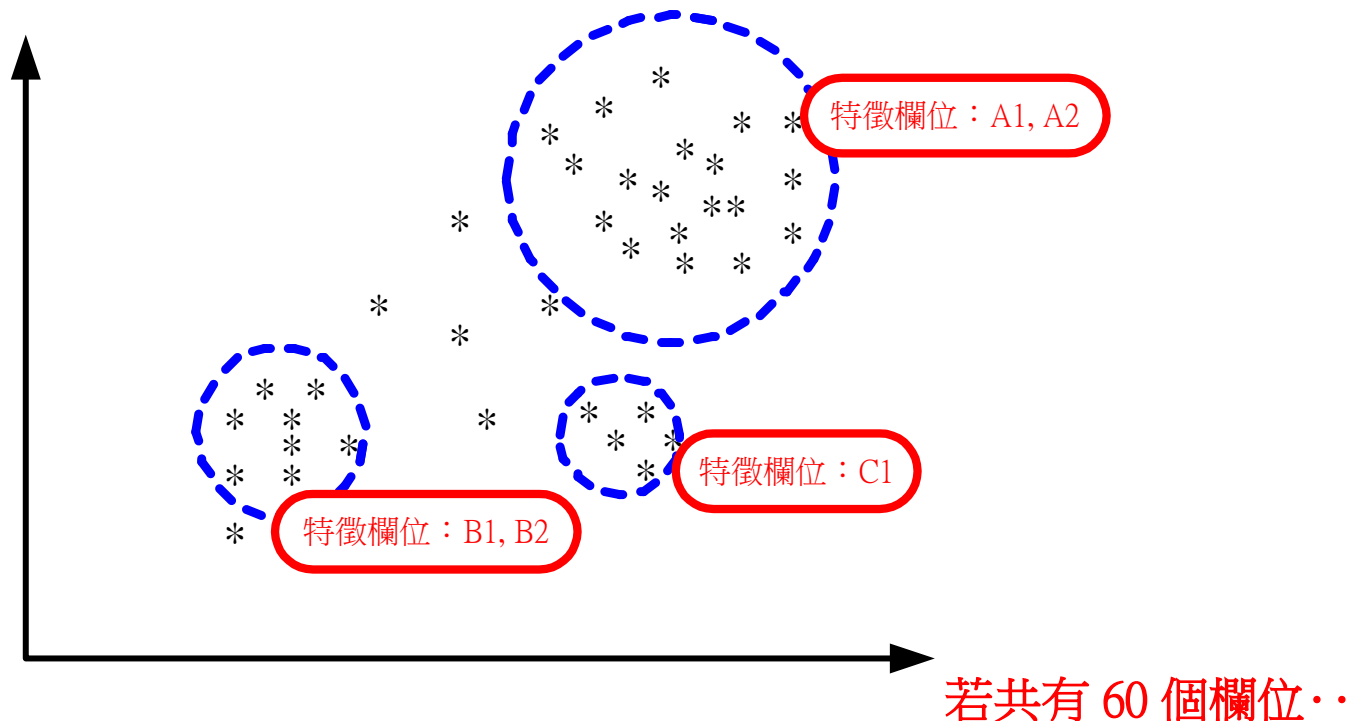
L₃

Itemset	Support
{2 3 5}	2

產品組合 {2,5} 或 {2,3,5}
最常被一起購買

1. 群集分析 Clustering

- ◆ 自動尋找大量欄位與資料中最主要與最顯著之群集分布
- ◆ 常見應用：從客戶中找出顯著之主力客群，做為目標市場區隔



2. 分類預測 Classification

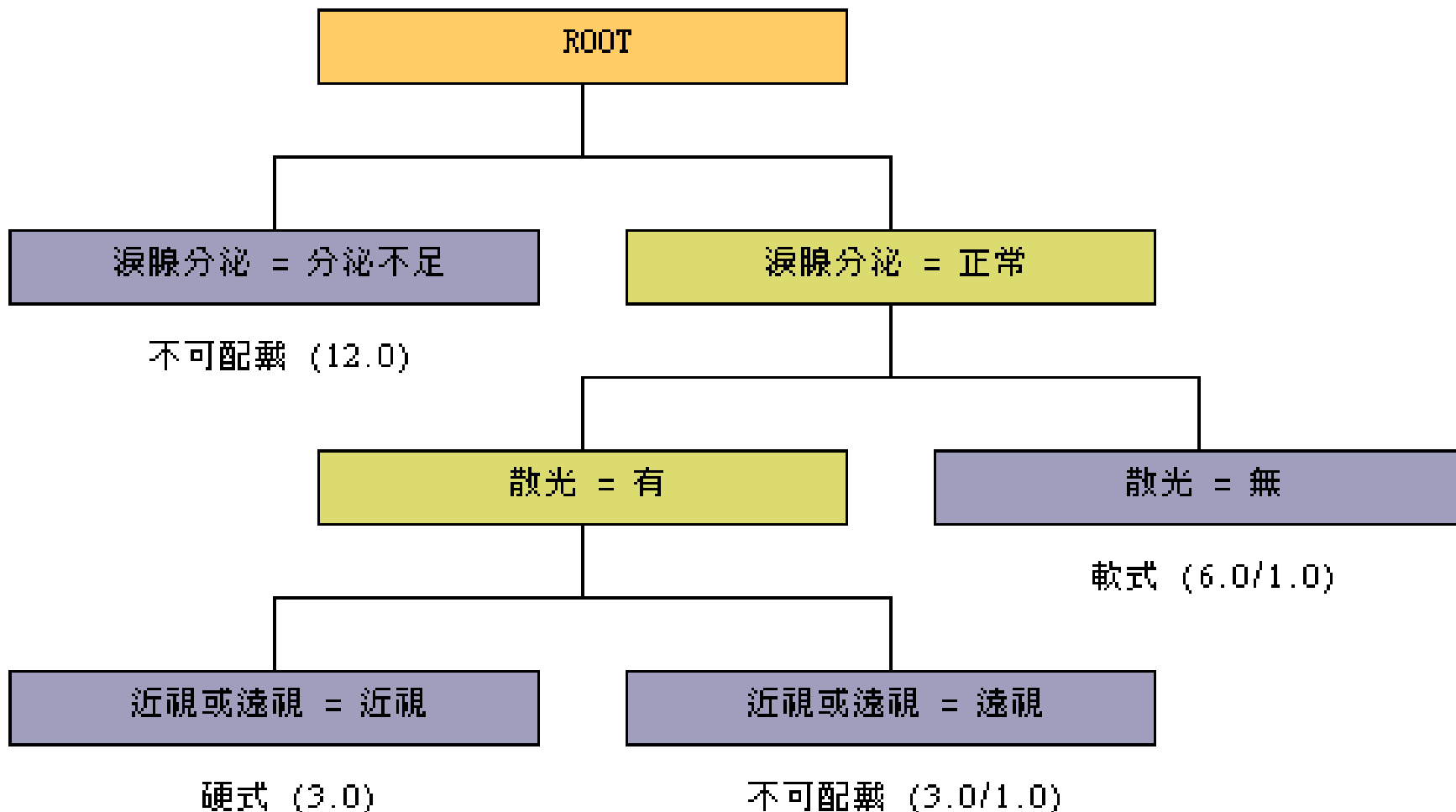
- ◆ 利用資料庫內每筆資料的已知欄位，預測目標欄位之值，並做為分類的依據
 - 可以將大量資料轉化成人類易於了解的知識樹
 - 常見應用：信用評等、消費行為預測、病症診斷

分類預測：眼科診所病例

年齡	近視或遠視	散光	淚腺分泌	隱形眼鏡處方
兒童	近視	無	分泌不足	不可配戴
兒童	近視	無	正常	軟式
兒童	近視	有	分泌不足	不可配戴
兒童	近視	有	正常	硬式
兒童	遠視	無	分泌不足	不可配戴
兒童	遠視	無	正常	軟式
兒童	遠視	有	分泌不足	不可配戴
兒童	遠視	有	正常	硬式
成人	近視	無	分泌不足	不可配戴
成人	近視	無	正常	軟式
成人	近視	有	分泌不足	不可配戴
成人	近視	有	正常	硬式
成人	遠視	無	分泌不足	不可配戴
成人	遠視	無	正常	軟式
成人	遠視	有	分泌不足	不可配戴
成人	遠視	有	正常	不可配戴
老年	近視	無	分泌不足	不可配戴
老年	近視	無	正常	不可配戴
老年	近視	有	分泌不足	不可配戴
老年	近視	有	正常	硬式
老年	遠視	無	分泌不足	不可配戴
老年	遠視	無	正常	軟式
老年	遠視	有	分泌不足	不可配戴
老年	遠視	有	正常	不可配戴

分類預測：眼科診所病例 (續)

- ◆ 自動選擇最佳分支條件，產生決策樹



3. 關聯規則 Association rules

- ◆ 尋找每筆交易中被同時購買之商品的關聯性

Buy (milk) → Buy (bread) 信心度 80 %

- ◆ 尋找消費者與商品之間關聯性

Nokia N95 → 男性、上班族、年收入80-120萬 信心度 60 %

- ◆ 亦可尋找任何人、事、物彼此間同時出現之關聯性

Support and Confidence

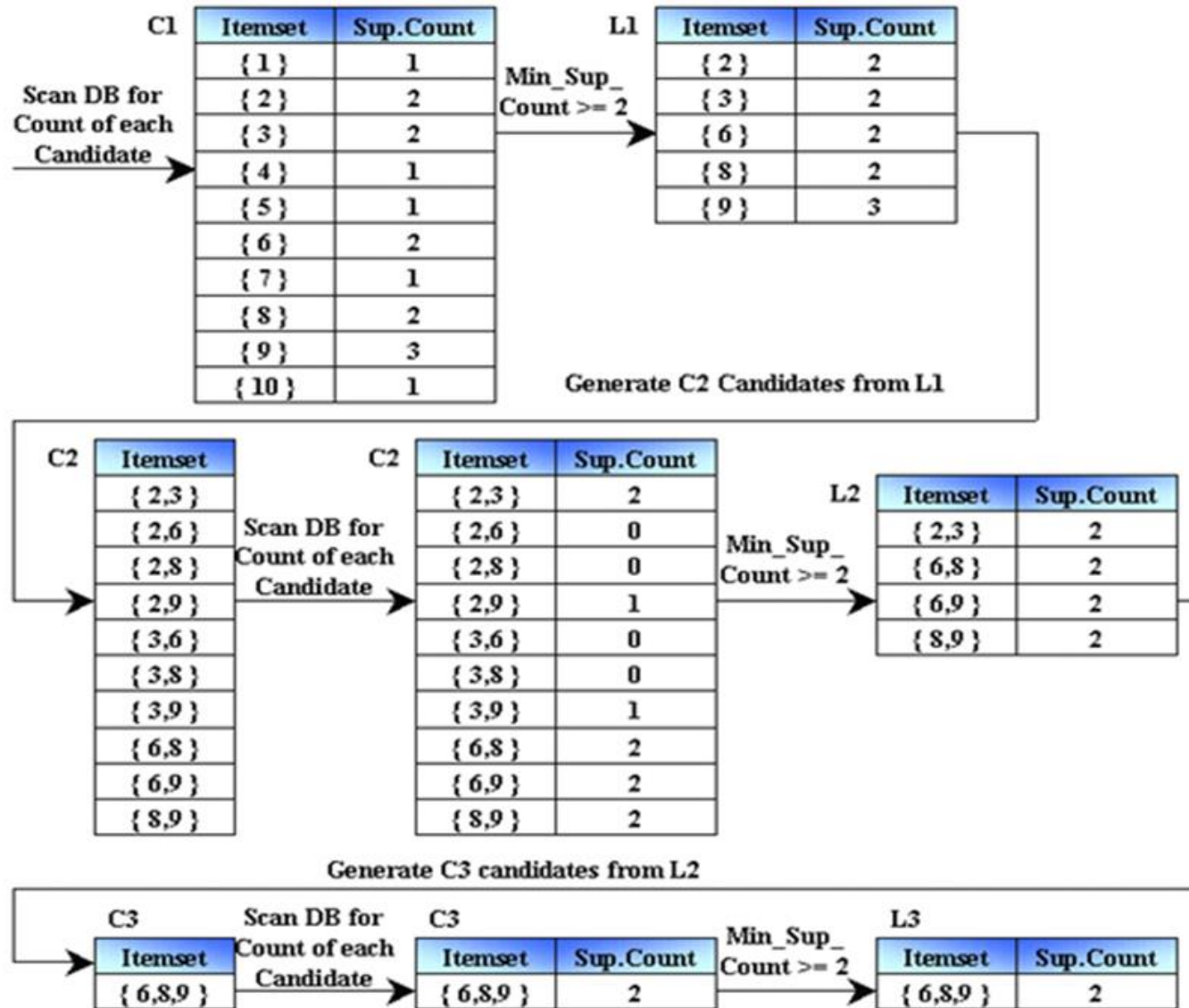
$$\text{支持度}(X \rightarrow Y) = \frac{\text{包含 X 和 Y 的交易紀錄}}{\text{總交易紀錄}}$$

$$\text{信心度}(X \rightarrow Y) = \frac{\text{包含 X 和 Y 的交易紀錄}}{\text{包含 X 的交易紀錄}}$$

Apriori

演算法

可以Java改寫
於Hadoop上
(MapReduce)



4. 連續行為 Sequential pattern

- ◆ 客戶購買某產品後之某段期間內，會再購買的產品

例：錄影帶

Star War → Empire Strikes Back → Return of the Jedi

- ◆ 常見應用：

- 消費者之消費行為預測
- 產品銷售預測
- 產品製程與存貨預測

連續行為 Sequential pattern (續)

客戶	日期	商品
John	19960701	Jurassic Park
John	19960712	Terminator 2 : Judgment Day
Bob	19960703	The Lion King
Bob	19960703	Sleepless in Seatle
Bob	19960712	Jurassic Park
Bob	19960723	Toy Story
Bob	19960723	Die Hard 3
Bob	19960723	Jurassic Park 2 : Lost World
Mary	19960708	Jurassic Park
Mary	19960708	Pretty Woman
Mary	19960708	Jurassic Park 2 : Lost World
Peter	19960706	Jurassic Park
Peter	19960713	Toy Story
Peter	19960713	Jurassic Park 2 : Lost World
Peter	19960715	Terminator 2 : Judgment Day
Nick	19960711	Terminator 2 : Judgment Day

連續行為 Sequential pattern (續)

◆ 最熱門連續行為

Jurassic Park → Toy Story , Jurassic Park 2 : Lost World

Jurassic Park → Terminator 2 : Judgment Day

◆ 行銷建議

- 產品合購優惠方案
- 櫃台人員主動推薦
- 內部商品擺設建議

案例 (1) 台灣最大實體書店

- ◆ 台灣地區大型書籍零售賣場領導品牌，擁有數十萬會員資料，每年會員交易紀錄超過數百萬筆
- ◆ 分析目標
 - 目標 1：尋找會員購買商品之間的關聯性
 - 目標 2：尋找會員基本資料、與購買商品之間的關聯性
- ◆ 樣本資料
 - 20萬筆會員資料
 - 10萬筆行銷活動收集之名單
 - 二年度的會員交易資料明細

台灣最大實體書店 (續)

- ◆ 針對**目標 1**，使用關聯分析 (Association) 模組，自動尋找出最具關聯性的購買商品
- ◆ 發現：
 - 購買 **休閒娛樂** 類商品的會員, 同時會再購買 **旅遊** 類商品
 - 購買 **乾隆相關** 書籍商品的會員, 同時會再購買 **雍正王朝 DVD**
- ◆ 意義：
 - 可以針對上述具高度關聯性的商品進行搭售與聯合促銷
 - 可以寄送另一商品之促銷訊息予只購買單一商品之會員
 - 賣場動線設計：具高度相關之商品應陳列在同一鄰近區域

台灣最大實體書店 (續)

- ◆ 針對**目標 2**，使用主力客群 (Clustering) 模組，自動尋找出會員資料中與商品特性關聯性最高的欄位
- ◆ 發現：
 - **旅遊** 類商品與會員資料中的 **性別** 與 **年齡** 欄位有高關聯性
 - 顯著區間：(Female, 30~40)
 - **財經** 類商品與會員資料中的 **職業** 與 **收入水準** 欄位有高關聯性
 - 顯著區間：(Employee, 500K~800K yearly)
- ◆ 意義：
 - **Direct Marketing**：可以將促銷商品 DM 只寄給最具關聯性的潛在客戶。可大幅降低行銷成本，並提高回應率與成交率

台灣最大實體書店 (續)

◆ 專案導入：

- 該專案執行期間, 由總經理指派專案小組負責
- 部份商品為少量多樣，如建築類、藝術類、國外進口書等
- 原先每年寄送的會員 DM 約 100 萬封，每封 DM 成本約 10-12 元，但平均回應率低於 2 %

◆ 效果：

- 經過資料分析後，了解客群分布，可進行精準的目標行銷
- 每年寄送的會員 DM 降為 20 萬封，回應率提高為 8-10 %
- 可以更準確地開發新客群，以及進行存書控制

案例 (2) 亞洲最大線上紅利集點網站

- ◆ 亞洲區最大的線上紅利集點網站，在四個區域 (九個國家) 中會員超過350萬人，專門經營線上紅利集點
- ◆ 分析目標
 - 目標 1：找出具有高紅利點數的會員族群之特徵
 - 目標 2：找出會員資料庫中的主要顯著客群
- ◆ 樣本資料
 - 350萬筆大中國區會員資料庫
 - 每位會員共有25個資料欄位，包括 ID, account, points, email, sex, nickname, birthday, join_date, job 等

線上紅利集點網站 (續)

- ◆ 針對**目標 1**，使用關聯分析 (Association)，自動尋找出與點數欄位最相關之欄位組合
- ◆ 發現
 - 紅利點數多的會員，與**地址Address**欄位高度相關
 - 住在**台北Taipei**的人點數最多 (Confidence 63.5%)
- ◆ 意義
 - 台北市的會員對於該線上紅利集點網站之大中國區盈收貢獻度最高，最有價值
 - 所有的會員經營與行銷預算，應集中火力於台北地區的會員

線上紅利集點網站 (續)

- ◆ 針對**目標 2**，使用群集分析 (Clustering)，自動尋找出特徵最集中的主要會員群集
- ◆ 發現：該網站350萬會員中，有三個最顯著的主力客群
 - **Group 1** (55.78%) : 年輕人、學生、男性
 - **Group 2** (19.53%) : 科技業、技術人員、工程師、或經理、男性、半數住在台北
 - **Group 3** (7.05%) : 科技及服務業、行政助理、秘書、服務人員、女性、半數住在台北
- ◆ 意義
 - 上述族群已超過 80%。可對這三個市場區隔，做行銷預算分配，進行更精準的直效行銷，將可大幅節省行銷成本，並提高效果。

案例 (3) 其它

◆ 某大電信業者：電信預警系統

- 依 Peak alarm (突來的不正常之尖峰用量) 與 Region alarm (連續罕用受話號碼或付費號碼) 判斷盜打行為
- 依系統警示信號預測系統擁塞機率，以預先調配郊區之基地台來支援，達成整體系統使用率之最佳化

在各產業的應用 – 以金融保險業為例

企業擁有許多的資料庫....

◆ 保戶基本資料

- 要保人：郵遞區號、生日、性別、婚姻等
- 被保人：郵遞區號、生日、性別、婚姻等

◆ 保單基本資料

- 險種代碼、保險金額、繳費方式、紅利發放方式等

◆ 保單繳費紀錄

- 保單年度、保單狀態、年度化保費、實繳保費等
- 業務單位、銷售管道、卡別等

企業的問題是....

- ◆ 這些存在 Excel 或 Access 的現成資料，能否協助
 - 了解既有客戶行為模式，做好客戶關係管理
 - 開發新客戶
 - 進行決策支援，選定目標市場與行銷策略
 - 降低行銷成本，提高回應率與成交率
 - 提高營收和顧客滿意度
- ◆ 如何可以達到上述目標？

1. 尋找保戶購買保單的決策模型

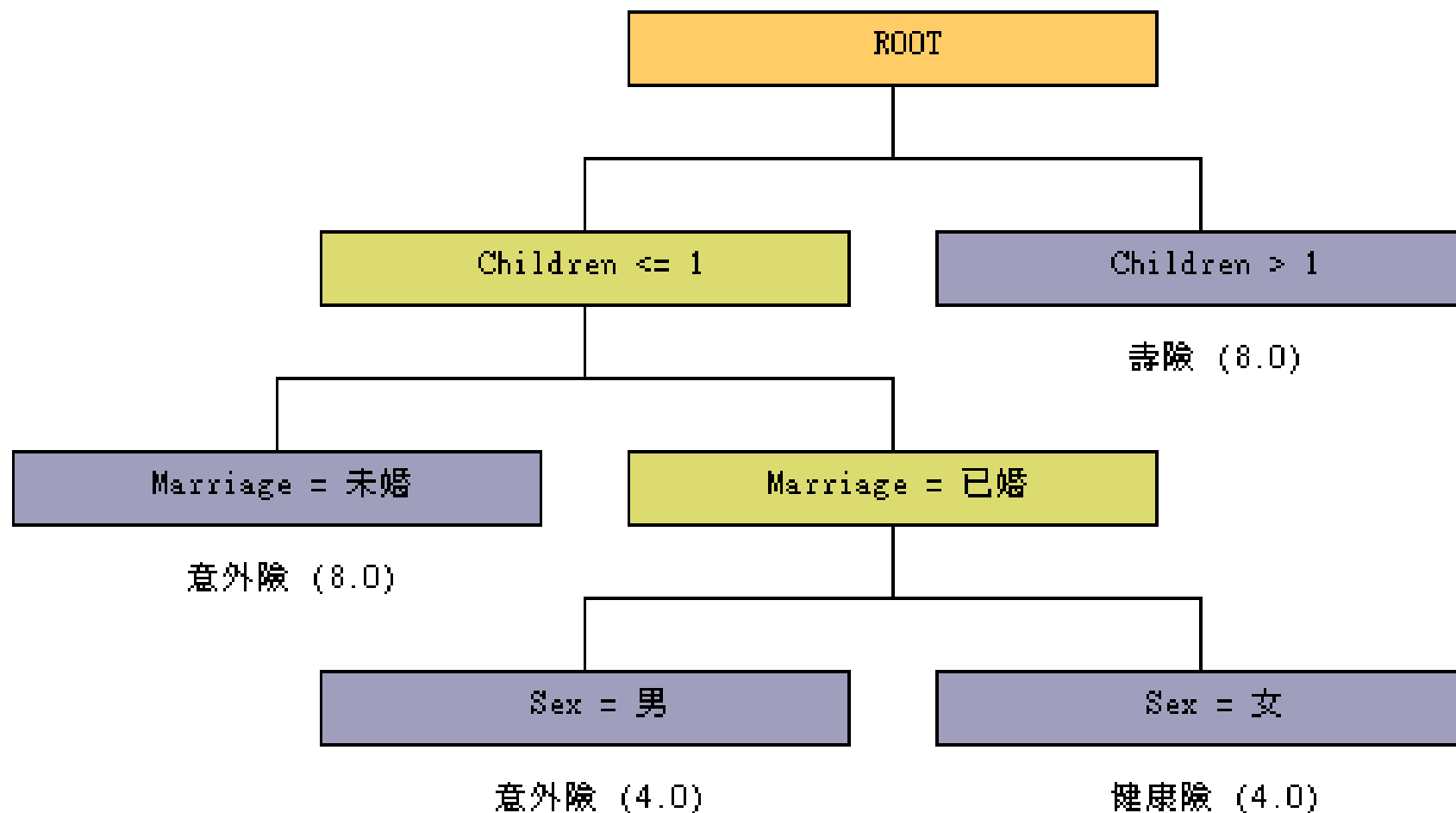
Sex	Age	Marriage	Children	Bonus	Product
男	22	未婚	0	無紅利	意外險
男	28	未婚	0	無紅利	意外險
男	38	未婚	1	現金	意外險
男	36	未婚	0	現金	意外險
女	24	未婚	0	抵繳保費	意外險
女	29	未婚	0	抵繳保費	意外險
女	24	未婚	0	無紅利	意外險
女	29	未婚	0	無紅利	意外險
女	35	已婚	0	無紅利	健康險
女	38	已婚	1	無紅利	健康險
男	47	已婚	2	現金	壽險
男	45	已婚	3	現金	壽險
男	32	已婚	0	抵繳保費	意外險
男	34	已婚	0	抵繳保費	意外險
女	37	已婚	0	現金	健康險
女	38	已婚	0	現金	健康險
女	46	已婚	2	抵繳保費	壽險
女	48	已婚	2	抵繳保費	壽險
女	46	已婚	2	無紅利	壽險
女	43	已婚	2	無紅利	壽險
男	36	已婚	0	無紅利	意外險
男	33	已婚	0	無紅利	意外險
女	44	已婚	3	無紅利	壽險
女	41	已婚	3	無紅利	壽險

保戶保單資料庫

- ◆ 要保人性別
- ◆ 要保人年齡
- ◆ 要保人婚姻狀況
- ◆ 要保人子女數
- ◆ 紅利發放方式
- ◆ 險種

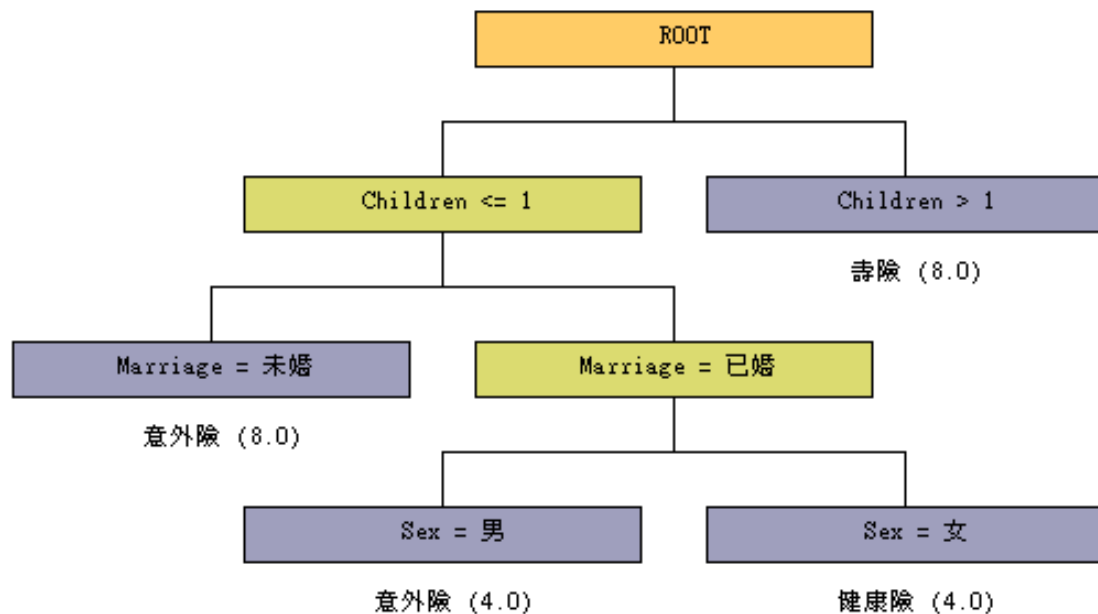
各險種之購買因素？

◆ 自動選擇最佳分支條件，產生決策樹



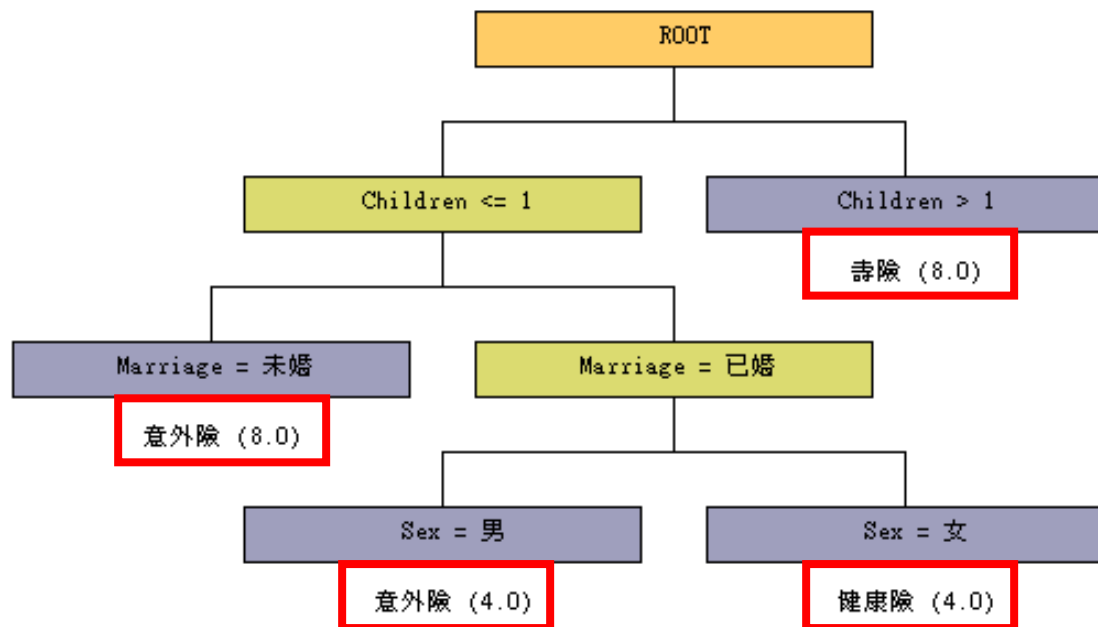
分析：建立客戶決策模型

Sex	Age	Marriage	Children	Bonus	Product
男	22	未婚	0	無紅利	意外險
男	28	未婚	0	無紅利	意外險
男	38	未婚	1	現金	意外險
男	36	未婚	0	現金	意外險
女	24	未婚	0	抵繳保費	意外險
女	29	未婚	0	抵繳保費	意外險
女	24	未婚	0	無紅利	意外險
女	29	未婚	0	無紅利	意外險
女	35	已婚	0	無紅利	健康險
女	38	已婚	1	無紅利	健康險
男	47	已婚	2	現金	壽險
男	45	已婚	3	現金	壽險
男	32	已婚	0	抵繳保費	意外險
男	34	已婚	0	抵繳保費	意外險
女	37	已婚	0	現金	健康險
女	38	已婚	0	現金	健康險
女	46	已婚	2	抵繳保費	壽險
女	48	已婚	2	抵繳保費	壽險
女	46	已婚	2	無紅利	壽險
女	43	已婚	2	無紅利	壽險
男	36	已婚	0	無紅利	意外險
男	33	已婚	0	無紅利	意外險
女	44	已婚	3	無紅利	壽險
女	41	已婚	3	無紅利	壽險



1. 自動嘗試所有欄位排列組合
2. 找出關鍵決策因素之優先順序
3. 自動切割適當值
4. 自動排除無關因素

應用：預測客戶行為



1. 依客戶狀況提出適當建議 透過Web或Mobile
2. 準確地大量開發新客戶 精準目標與擴散
3. 「科學化」降低行銷成本，提高成交率和營收

2. 尋找最熱門之保戶保單的關聯性

◆ 分析目標

- 目標 1 : 保戶基本資料、和購買保單間的熱門關聯性
- 目標 2 : 購買保單之間的關聯性

◆ 分析結果

(新鴻運終身壽險, 300-400萬) → (要保人: M, 被保人: F)

(萬福增額終身壽險) → (要保人: F, 30-40歲, 被保人: F, 0-10歲)

(新鴻運終身壽險) → (防癌保本終身健康保險)

◆ 如何應用熱門關聯性？

(新鴻運終身壽險, 300-400萬) → (要保人: M, 被保人: F)

(萬福增額終身壽險) → (要保人: F, 30-40歲, 被保人: F, 0-10歲)

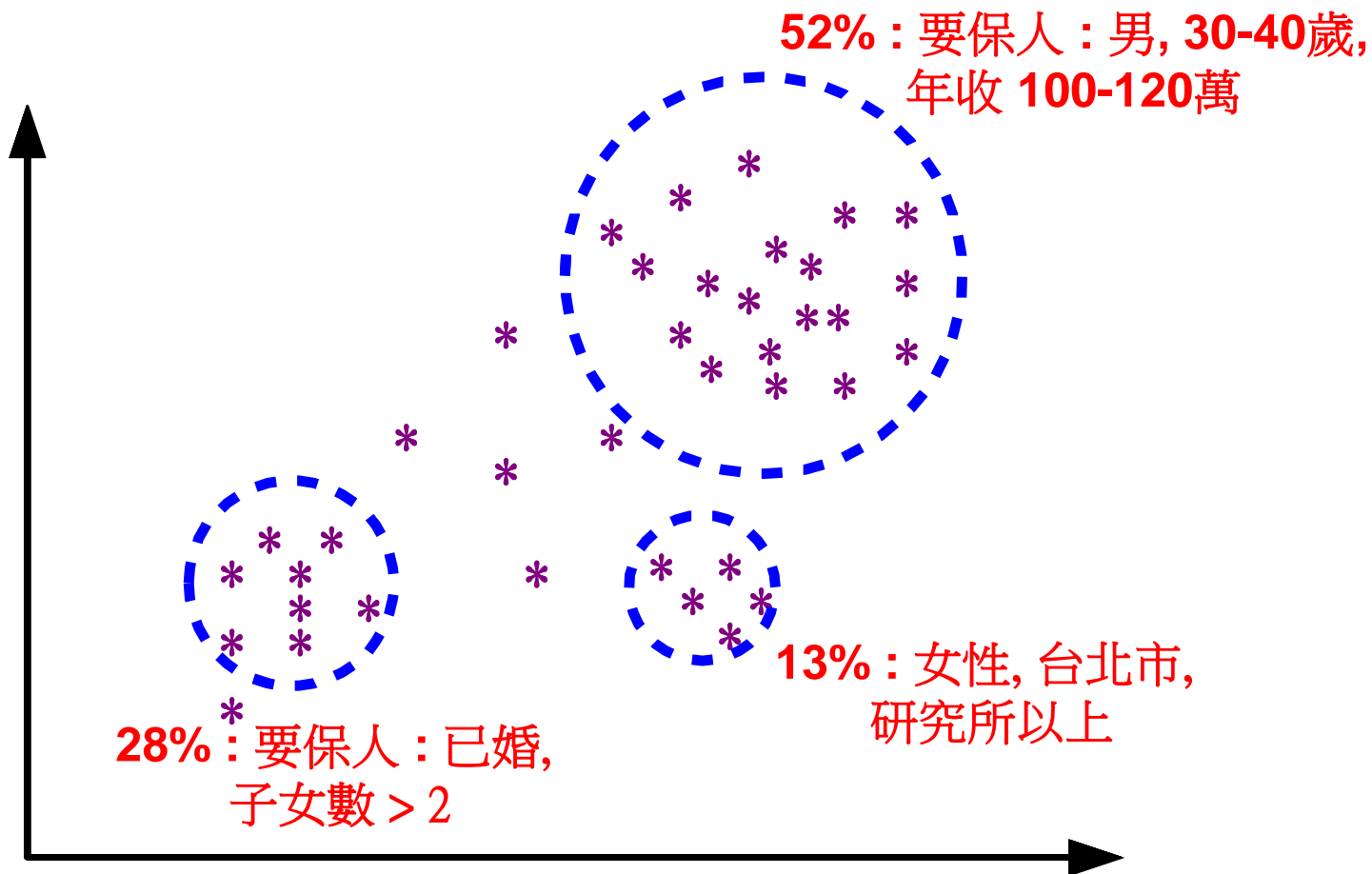
- **直效行銷**：可以將商品資訊只推銷給最具關聯性的潛在客戶。可大幅降低行銷成本，並提高回應率與成交率

(新鴻運終身壽險) → (防癌保本終身健康保險)

- **搭售、聯合促銷、交叉行銷**：可以寄送另一商品之促銷訊息予只購買單一商品之會員，創造新的營收

3. 尋找主力保戶客群之特徵

- ◆ 盈收貢獻度問題
- ◆ 想了解購買三張保單，或累計投保1000萬以上的主力客群特徵？



◆ 如何應用主力客群特徵分析？

52%：要保人：男, 30-40歲, 年收 100-120萬

28%：要保人：已婚, 子女數 > 2

13%：女性, 台北市, 研究所以上

1. 行銷預算之分配依據

2. 依特徵不同訂作行銷策略

3. 喚醒具同樣特徵但貢獻度仍不高的潛力客群

符合（要保人：男, 30-40歲, 年收 100-120萬）

但目前（購買三張保單或累計投保1000萬以下）之族群

在各產業的應用 – 以零售通路為例

目標 1 聯合促銷

STEP 1 : 目標

- 收集購買紀錄中，分析商品之間的關聯性

STEP 2 : 分析

- 可依不同的資料尺度進行分析

部門 - 生鮮食品, 摩登服飾, 3C家電等

專櫃 - 佳麗寶, 資生堂, 美爽爽等

品群 - 化妝品, 皮件, 女鞋等

商品 - 水亮唇膏

- 分析結果

中型免洗盤 → 竹籤 and 竹筷

信心指數 74.3 %

車用吸塵器 → 車用垃圾筒

信心指數 42.8 %

佳麗寶 PN 化妝品 → 華歌爾無肩帶胸罩

信心指數 65.2 %

目標 1 聯合促銷 (續)

STEP 3 : 執行

- 所代表的意義

- 商品特性關聯：車用吸塵器 → 車用垃圾筒 (清潔車內)
- 消費目的關聯：中型免洗盤 → 竹籤 and 竹筷 (郊遊烤肉)
- 消費族群關聯：佳麗寶 PN 化妝品 → 華歌爾無肩帶胸罩 (年輕流行)

- 擬定促銷策略

- 搭售
- 聯合促銷：主題特賣會、折價券
- 亦可做為商場動線設計，以及進補貨之參考

目標 2 直效行銷

STEP 1 : 目標

- 從會員資料與購買紀錄中，找出會員與商品之間的關聯性
- 從會員資料與購買紀錄中，尋找會員的連續購買行為

STEP 2 : 分析

- 分析結果

果汁飲料類 → 女性, 25-35歲, 台北縣市	信心指數 63.2 %
汽車百貨 → 男性, 25-40歲	信心指數 74.1 %
愛狗座墊 → 項圈 → 狗食 or 除蚤劑	信心指數 59.8 %
變速電鑽 → 捲尺 → 多功能噴漆	信心指數 45.3 %

目標 2 直效行銷 (續)

STEP 3 : 執行

- 擬定直效行銷策略

- 針對目標族群寄送行銷訊息 → 成本更低，回應率更高

- 果汁飲料類 → 女性, 25-35歲 信心指數 63.2 %

- 汽車百貨 → 男性, 25-40歲, 台北縣市 信心指數 74.1 %

- 預測消費者購買行為，刺激消費 → 增加營收

- 愛狗座墊 → 項圈 → 狗食 or 除蚤劑 信心指數 59.8 %

- 變速電鑽 → 捲尺 → 多功能噴漆 信心指數 45.3 %

- 辨識目標族群，加以行銷擴散

- 建立「愛狗族群張貼可愛照片」，或「DIY家居族群分享經驗」

(2) 非結構性資料分析

◆ 轉結構性資料

- 命名實體擷取 Named entity extraction
- 斷詞分析 / 關鍵字詞擷取 Power term extraction
- 標記 Tagging

◆ 其它常見的 Text Mining 模組

- 相關性 Relevance
- 意見觀點及情緒 Opinion & Sentiment
- 群集分析 Clustering
- 分類預測 Classification
- 摘要 Summarization

巨量資料分析技術 (3) – Text Mining

Text Mining

- ◆ The **non-trivial** extraction of **implicit, previously unknown, and potentially useful** information from (large amount of) data – NCSA
- ◆ Information retrieval (search) is the foundation
 - " It's impossible to create knowledge from information that cannot be found or retrieved. " – IDC

Text Mining is not ...

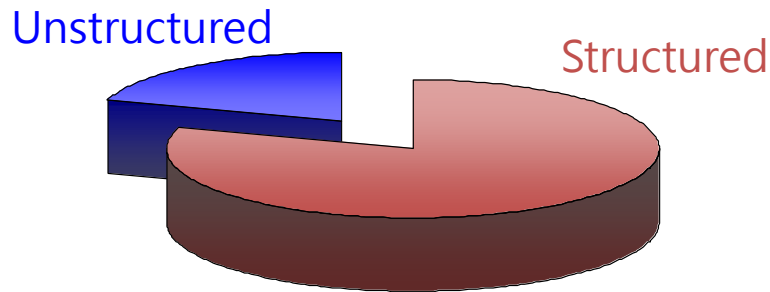
- ◆ Text understanding 意義理解
- ◆ Text composing 文字創作
- ◆ Machine translation 機器翻譯
- ◆ Rhetoric expression 修辭表達

Text mining for Content Intelligence

- ◆ 結構性資料 Structured
- ◆ 資料庫 Database
- ◆ 非結構性資料
- ◆ 檔案系統 File System
- ◆ 網站 Web Site
- ◆ 文件管理 Document Management
- ◆ 內容管理 Content Management
- ◆ 資料探勘 Data Mining
- ◆ 文字探勘 Text Mining
- ◆ 商業智慧 Business Intelligence
- ◆ 內容智慧 Content Intelligence

智慧資產的價值存在於非結構性資料

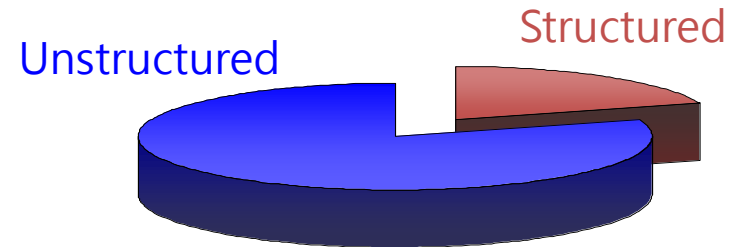
IT INVESTMENT



IT的投資大多集中在結構性資料

- ERP
- BI
- CRM
- SCM
- Accounting

BUSINESS VALUE



大多數對組織有幫助的智慧資產都存在於非結構性資料

- File Systems
- Web server
- Document Management
- Content Management

Source: Document and Content Technologies Market Forecast and Analysis, IDC

Content Intelligence Platform

使用者



決策者



業務



研發人員



合作伙伴



客戶

應用程式



決策系統



行動通訊



應用程式



企業入口網站



外部網站

解決方案

內容智慧平台

資料源



檔案系統



群組軟體



內容管理

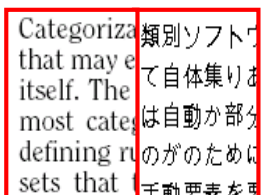


資料庫



網際網路

資料格式



文字

2003	2004	2005
382.0	536.4	738.0
256.7	358.6	492.0
341.0	471.4	650.1
501.3	736.8	1,019.1
284.4	405.6	562.3
66.3	95.5	131.8

數據



圖片



聲音



影音

命名實體標記

- ◆ 採用模式比對與機器學習演算法
 - Support Vector Machine (SVM)
 - Conditional Random Field (CRF)
 - 人名準確率可高達97%
 - 可學習，可調校，可設定黑白名單

蔡金愛 女 1962/**/14 48 無業 嘉義縣 Q20062**** 臺中市西區 中興里美村路一段***巷**號 詐欺 涉嫌詐欺 2011/**/31 臺中市南屯區 Y Z:\Photos\1000232蔡金愛涉嫌詐欺_1.JPG 楊正太 Y 1 Y Y 2 1 Y Y N N 余日天 男 1971/**/01 49 服務工作人員 南投縣 告訴代理人 臺中市西屯區福上巷**弄**號 詐欺 臺中市南屯區大墩路***號*樓 (家樂福) 臺中市南屯區大墩路***號*樓 一、犯罪嫌疑人蔡金愛於101年*月31日15時涉嫌詐欺罪犯嫌蔡金愛、無業，經查有賭博刑案紀錄，尤不知悛悔。涉嫌於上記犯罪時、地，意圖為自己不法之所有，進入告訴代理人余日天所任職之台灣家樂福股份有限公司大墩店內，利用該店放置麻將牌區域未裝設監視器監錄，徒手打開二盒價錢不同之麻將牌外盒膠膜，再將高

body h1 span b span

命名實體標記 – 流程

分析文件

堂潮公司 姓名：李世民
性別：1男0女
婚姻狀況：1未婚0已婚
年齡39歲 / 民國62年3月
21日出生
住址：台北市烏有街三段
十號五樓
電話：02-2712345
Email：smart@mail.tw
...



個資分析結果

[1] 李世民
Type:Name
Score:3.0
Index:7,436,500

[2] 堂潮公司
Type:Organization
Score:1.0
Index:0

[3] 台北市烏有街三段十號五樓
Type:Address
Score:1.0
Index:51

[4] 02-2712345
Type:Phone_num
Score:1.0
Index:65

斷詞分析 / 關鍵字詞擷取

- 地方新聞 ▶
- 兩岸台商 ▶
- 全球觀察 ▶
- 意見評論 ▶
- 財經產業 ▶
- 股市投資 ▶
- 基金理財 ▶
- 運動大聯盟 ▶
- 數位資訊 ▶
- 娛樂追星 ▶
- 消費流行 ▶
- 生活天氣 ▶
- 健康醫藥 ▶
- 旅遊休閒 ▶
- 校園博覽會 ▶
- 閱讀藝文 ▶
- 聯合書報攤 ▶
- 網路購物 ▶
- 數位閱讀 ▶
- 進修線上 ▶
- 職場行家 ▶

美10月消費者信心 41年來最低

【經濟日報/編譯莊雅婷/綜合彭博資訊二十八日電】

2008.10.29 02:28 am

股市重挫與信用市場凍結，拖累美國10月消費者信心指數創空前新低，美國20大城市8月房價也創下史上最大跌幅。

美國經濟研究聯合會28日表示，10月消費者信心指數跌至38，遠不及9月經調整後的61.4，也是1967年以來最低水準；分析師原先預估該指數會跌到52。

標準普爾500指數逼近70年最大單月跌幅，使許多家庭財富蒸發、房屋淨資產縮水，失業率也跟著攀升。經濟前景烏雲罩頂，暗示占國內生產毛額（GDP）比重三分之二的消費者支出恐進一步惡化，將令美國景氣雪上加霜。

花旗集團全球市場首席經濟學家狄克利蒙提說：「消費者信心崩盤似乎加重經濟頹勢，就業市場疲軟的狀況會更加嚴重。」

法拍屋數目激增，使8月標普／席勒房價指數比去年同期下跌16.6%，自2007年1月起，這項指標每月都下滑。

數據顯示，美國20大城市的房價年增率已連續第五個月下跌，其中鳳凰城和拉斯維加斯的房價都跌逾三成，狀況最慘。

[Nb 花旗] [Na 集團] [Nc 全球] [Nc 市場] [Na 首席] [Na 經濟學家] [Nb 狄克利蒙提] [VE 說]
[COLONCATEGORY :] [PARENTHESISCATEGORY 「] [Na 消費者] [Na 信心] [UH 崩盤] [D 似乎] [UC
加重] [Na 經濟] [Na 頹勢] [COMMACATEGORY ,] [Nu 就業] [Nc 市場] [UH 疲軟] [DE 的] [Na
狀況] [D 會] [Dfa 更加] [UH 嚴重] [PERIODCATEGORY 。] [PARENTHESISCATEGORY 」]

斷詞分析 / 關鍵字詞擷取 – 客服紀錄

- Tagging 人名、關鍵詞、時間、地點、情緒
- Summary 摘要
- Tracking 追蹤

SA memo	TEXT MINING 結果								後續應用
	註記	提醒	抱怨	偏好	理財	日期	摘要	關鍵字	
美金外幣定存資金不想讓先生知道, 因為是債券型基金配息累積的, 想當做私房錢	✓						美金外幣定存資金不想讓先生知道		特殊背景註記
日前聽聞FP說土地信託的詐騙方式, 有點擔心客戶, 但是因為不了解手法, 再調查, 11/20客戶會來再請他小心		✓				11/20	土地信託詐騙		下名單
客戶反應中信銀保險的客戶禮券太晚收到, 對中信銀保險印象不佳			✓				中信銀保險的客戶禮券太晚收到		即時pass至客經三部
客戶出國1個月帶回萬張攝影照片, 由秘書代為解檔處理. 喜號旅遊攝影. 下星期與PA再約訪客戶見面.				✓				旅遊, 攝影	特殊背景註記/ 鐵樂美

斷詞分析 / 關鍵字詞擷取 – 大量文件

- Thesaurus 語彙庫 – 以犯罪偵防為例

事件	物品
行車不慎、毆打、捏造、陪酒 踹開、遭竊、攻擊、侮辱、側 錄、竊得、擷圖、擄人、傷害 勒贖、偽造、掐住、飲用酒精 砸破、竊盜、順手拿走、押上 車、橫越車道、收取高額利息 誣告、恐嚇、塗改...	侮辱字眼、侮辱言詞、不實文 字、不實文書、不實內容、不 實言論、不雅文字、不實憑證 不實謠言、動物用藥、爆竹煙 火、成人貼圖、龐大利息、詐 騙教戰手冊、白鼻心、居留資 料、水車、印鑑章、支票...

標記與相關性分析

◆ 多種來源與資料格式

- 包括電子檔案與紙本掃描圖檔

◆ 進行資料整備、匯入、分析

◆ 個資實體與關係擷取

人名, 組織, 電話, 地址, 帳號, 車牌

時間, IP與email, 婚姻, 學經歷等

親屬, 同艙, 同事, 朋友, 同學, 同囚

通聯, 以及人事時地物間之關係

	人名	組織	電話	地址	帳號	車牌	時間
親屬	●			●			●
聯徵	●				●		●
健保在保	●	●	●	●			●
大陸人士	●			●			●
艙單	●			●			●
人際網路	●						●
貨物通關		●					●
入出境	●			●			●
Gmail等							●
戶役政相片	●			●			●
工商	●	●		●			●
車輛違規	●			●		●	●
ATM	●	●			●		●
金流	●	●		●			●
地政	●			●			●
護照簽證	●			●			●
同事	●	●		●			●
護照相片	●						●
Facebook							
健保就醫	●	●	●	●			●
員工投保 (健保)	●	●					●
投保歷程 (健保)	●	●		●			●
健保其他	●	●					●
車籍	●			●		●	●
Plurk							
行車記錄							

關聯分析模式

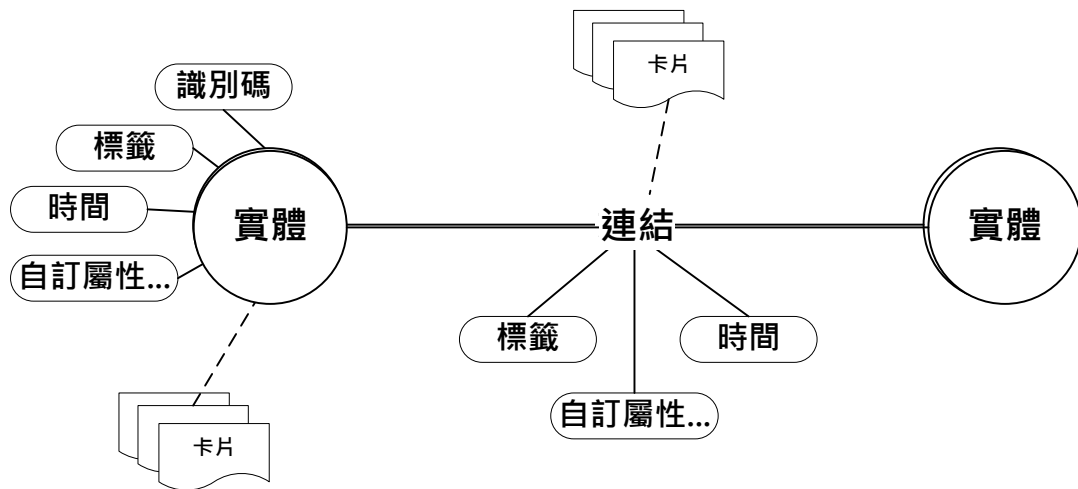
◆ 實體關聯

- 依實體進行關聯分析

Ex. Apriori

◆ 實體時間序列分析

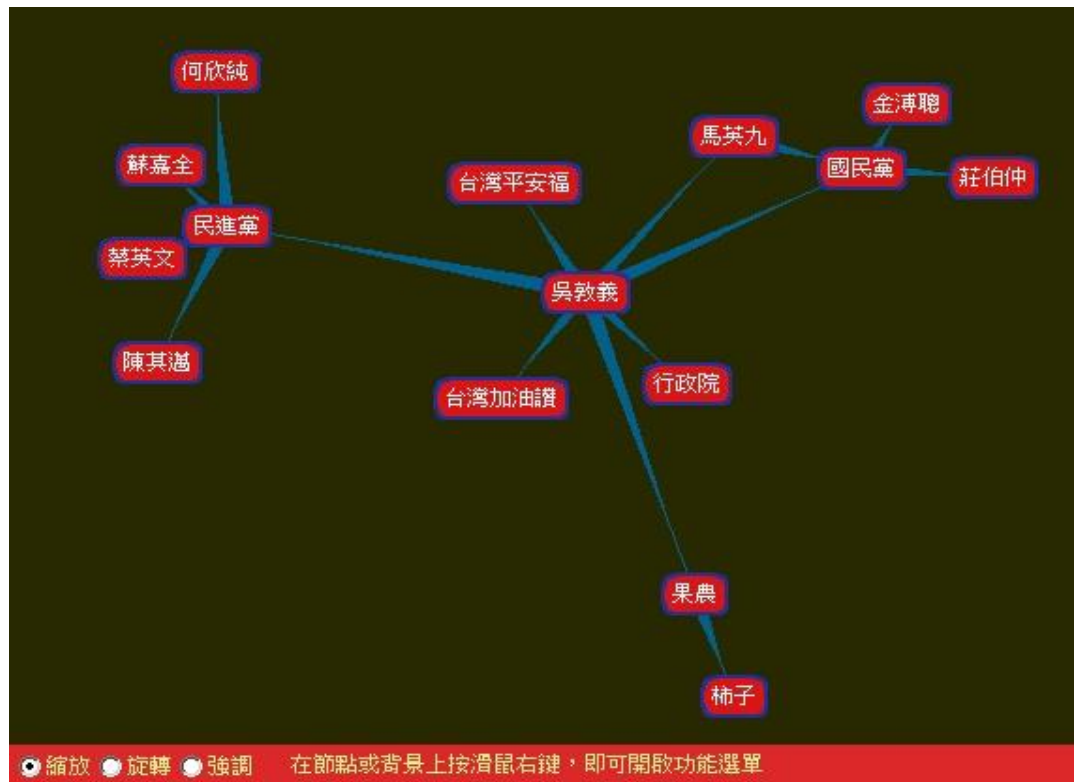
- 依特定的時間區間繪製成時間走勢圖
- Sequential pattern



關聯分析模式

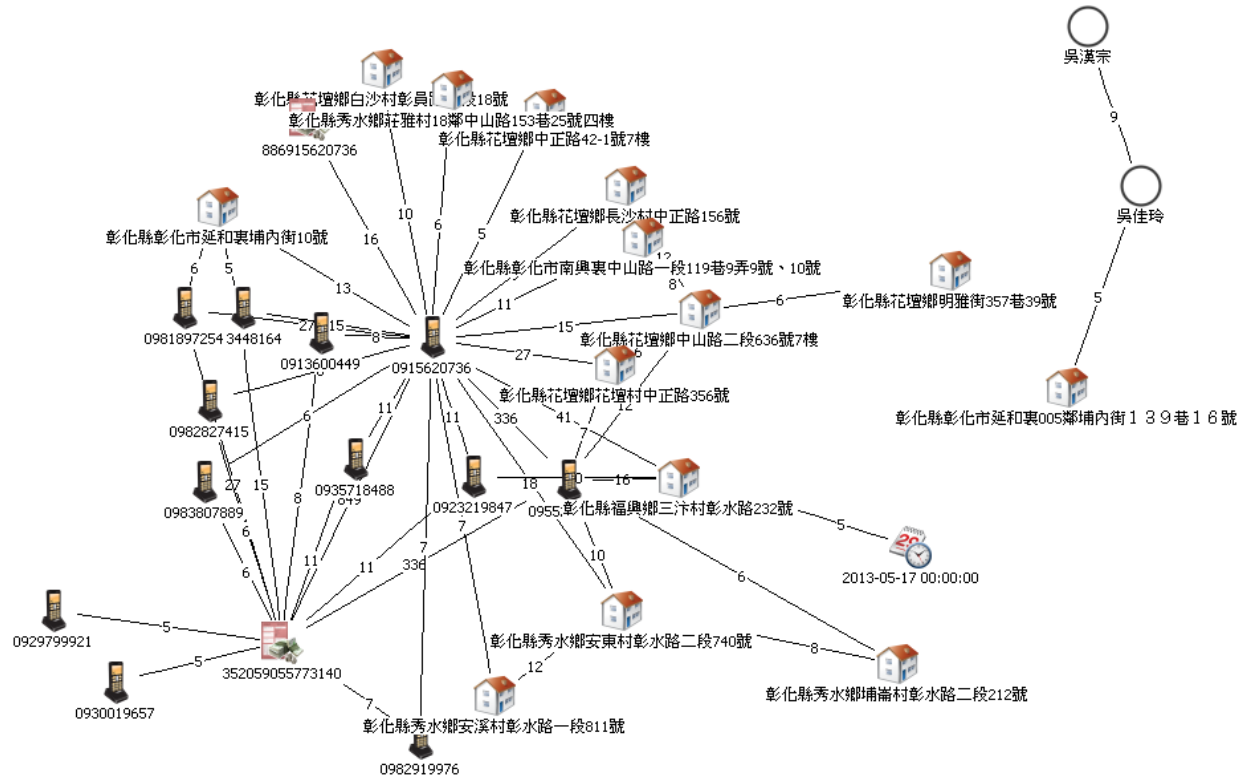
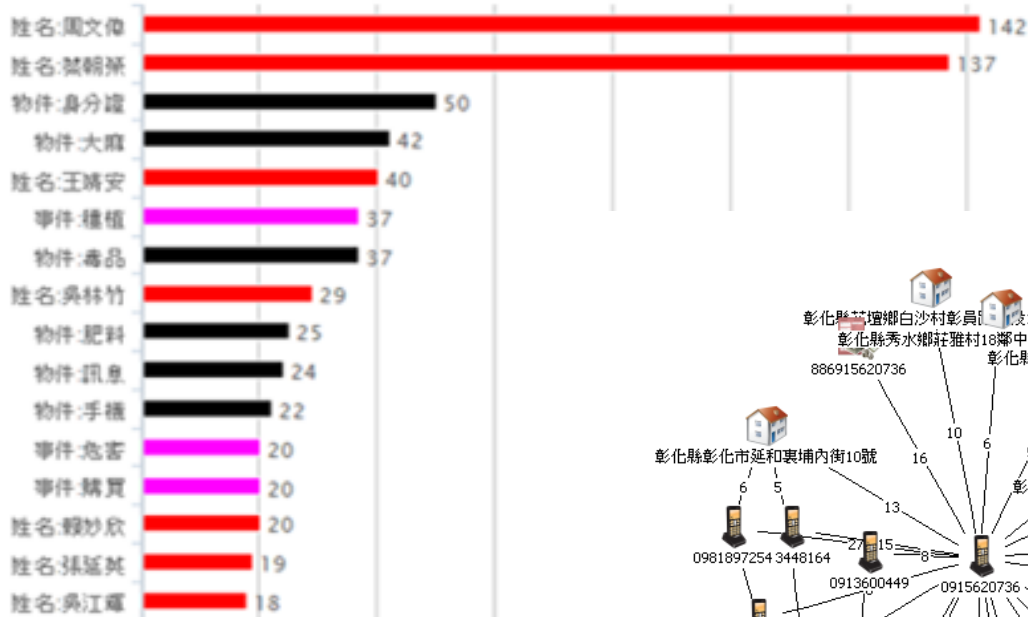
◆ 網絡分析

- 利用關聯分析製成網絡全圖
- 提供 expand & drill down

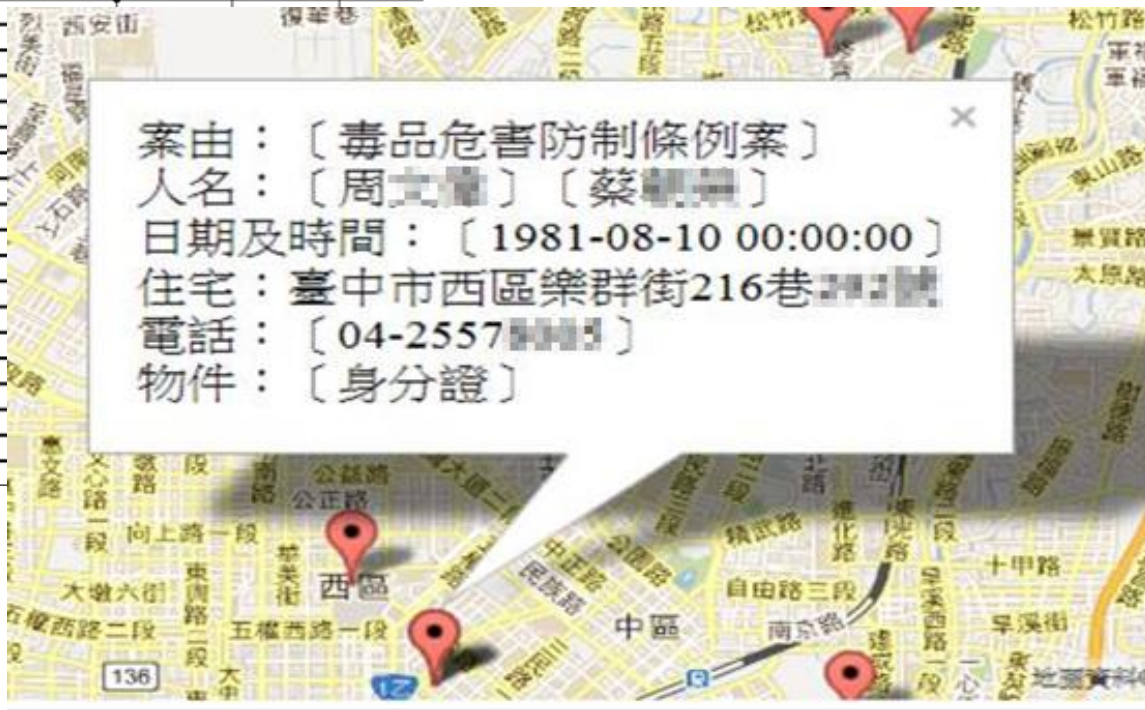


更多分析能力與視覺化 (1)

分析結果-熱點分析

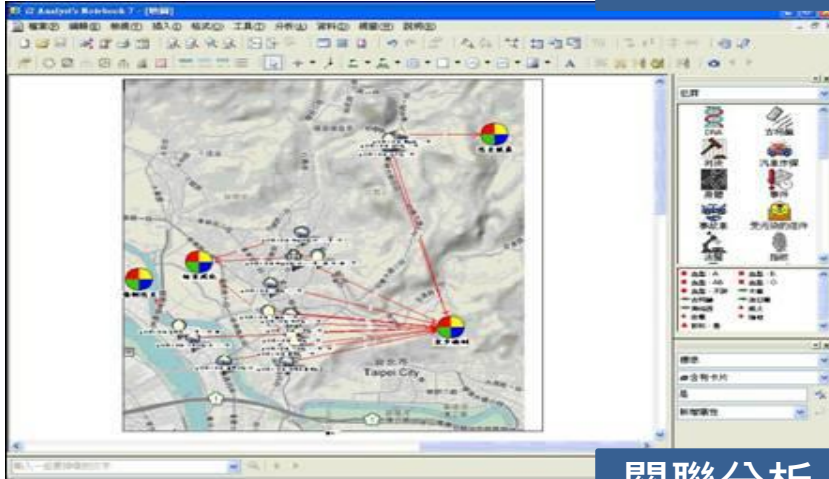


更多分析能力與視覺化 (2)

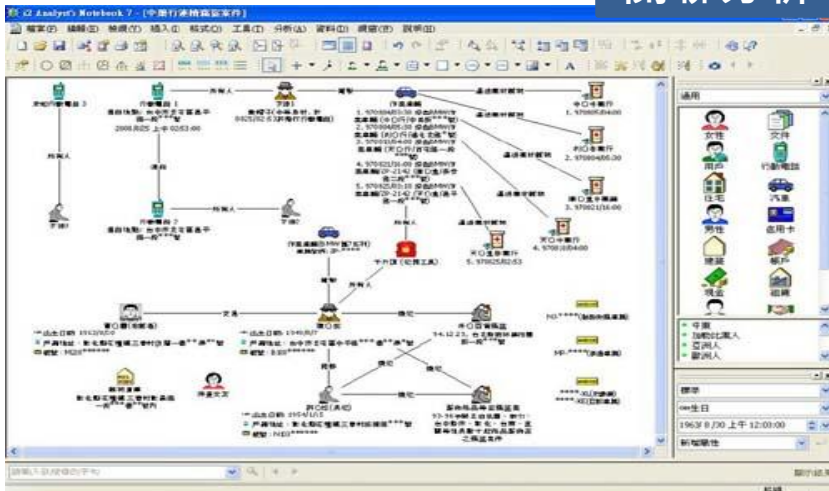


更多分析能力與視覺化 (3)

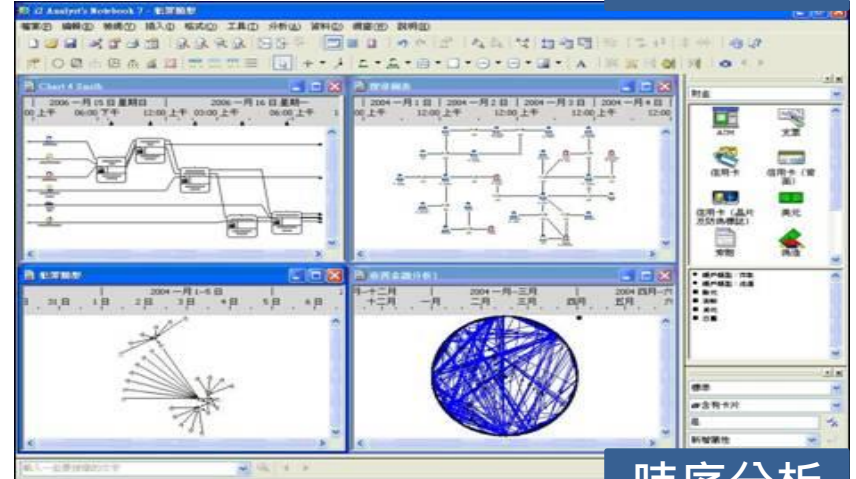
空間分析



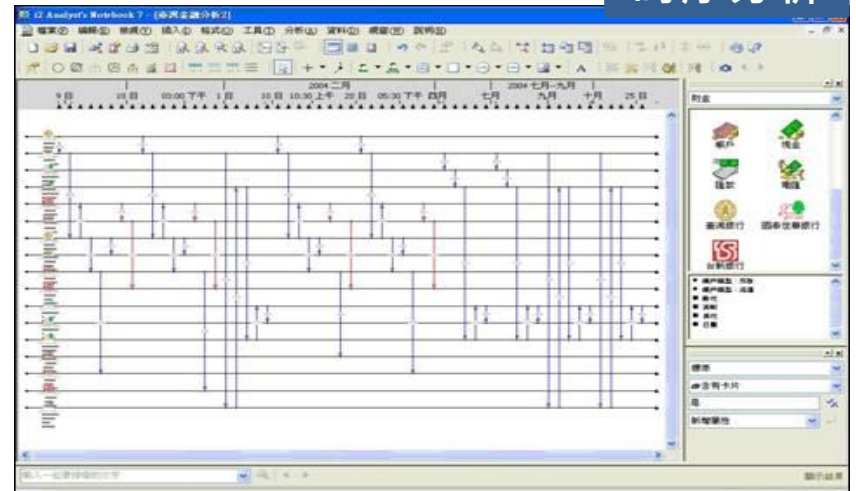
關聯分析



網絡分析



時序分析



個案研究：企業搜尋引擎與語意分析平台

Tornado Search / ENLP Platform

結果呈現

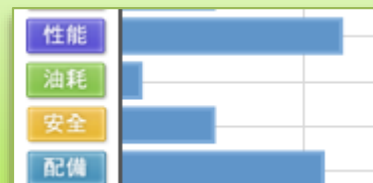
關聯圖 (知識地圖)



風暴圖



分類導覽



實體關聯

查詢輔助
智慧提示

形似相關

自動摘要

相關文件

情緒分析

意見評價

內容分析

關聯計算

分類引擎

語意解析

實體擷取

詞彙擷取

斷詞 / 新詞

地址

人名/組織名

電話

帳號

自訂詞庫

非結構資料
整合檢索

Tornado Search
Platform

龍捲風知識檢索平台



» 問題討論